a characteristic comparison routine identifying the file content as having a characteristic based on the appearance of the digital content ID in the appearance database.

2. The content classification system of claim 1 wherein said ID generator comprises a hashing algorithm.

3. The content classification system of claim 2 wherein said hashing algorithm is the MD5 hashing algorithm.

4. The content classification system of claim 1 wherein said ID appearance database tracks the frequency of appearance of a digital ID.

5. Please Cancel Claim 5.

A3

6. The content classification system of claim 1 wherein said plurality of digital ID generators are coupled to said database via a combination of public and private networks.

7. The content classification system of claim 6 wherein said database is coupled to an intermediate server which is coupled to said plurality of generators.

8. The content classification system of claim 6 wherein said intermediate server is a web server.

9. The content classification system of claim 1 wherein said characteristic comprises junk e-mail and said characteristic is defined by a frequency of appearance of a digital ID.

10. A method for identifying characteristics of data files, comprising:

receiving digital content identifiers for the data files from a plurality of source systems all coupled to a network;

determining, on a processing system coupled to the network, whether the forwarded identifier matches a characteristic of other identifiers; and

outputting, to at least one of the plurality of source systems responsive to a request from said source system, an indication of the chrematistic of the data file based on said step of determining.

11. The method of claim 9 wherein said step of generating comprises hashing at least a portion of the data file.

12. The method of claim 10 wherein said step of hashing comprises using the MD5 hash.

13. The method of claim 10 wherein said step of generating comprises hashing multiple portions of the data file.

A3

14. The method of claim 9 wherein said data file is an email message and said step of determining comprises determining whether said email is SPAM.

15. The method of claim 9 wherein said step of determining identifies said e-mail as SPAM by tracking the rate per unit time a digital ID is generated.

16. Please cancel claim 16.

17. The method of claim 15 wherein said step of processing comprises instructing said plurality of source systems to perform an action with the email based on said determining step.

18. A method of filtering an email message, comprising:

receiving a digital content identifier unique to the message content from at least two of a plurality of devices;

9

comparing the digital identifier to a characteristic database of digital identifiers received from said plurality of devices to determine whether the message has a characteristic; and

responding to a query from at least one of said plurality of devices of the existence or absence of said characteristic of the message based on said comparing.

19. The method of claim 17 wherein said step of comparing occurs on at least one network coupled processing system.

20. The method of claim 18 wherein said step of receiving includes receiving identifiers from said plurality of first systems.

21. The method of claim 18 wherein said plurality of systems are coupled by the Internet.

A3

22. The method of claim 18 wherein said step of comparing comprises determining the frequency of a particular ID occurring in a time period, classifying said ID as having a characteristic, and comparing digital identifiers to said classified IDs.

23. A file content classification system, comprising:

a first system having a file to be classified;

a file ID generator on the first system outputting at least one file ID for the file based on a generated checksum of at least one selected portion of said file;

a database on a second system coupled to the ID generator to receive IDs generated by the ID generator; and

a comparison routine on the second system classifying the ID relative to the database as meeting or not meeting a characteristic.

24. The system of claim 22 including a plurality of first systems each including a respective file ID generator coupled to the database on the second system.

9

25. The system of claim 23 wherein the plurality of first systems is coupled to the second system via a combination of public and private networks.

26. The system of claim 24 wherein the second system comprises a web server interface system and a database system, wherein the database system is isolated from the Internet by the web server system.

27. A file content classification system for a first and second computer coupled by a network, comprising:

a client agent file content identifier generator on the first computer, the file content identifier comprising a checksum of at least two non-contiguous sections of data in a file; and

a server comparison agent and data structure on the second computer receiving identifiers from the client agent and providing replies to the client agent;

wherein the client agent processes the file based on replies from the server comparison agent.

A3

28. A method for providing a service on the Internet, comprising:

collecting data from a plurality of systems having a client agent generating digital content identifiers for each of a plurality of files on the Internet to a server having a database;

characterizing the files based on said digital content identifiers received relative to other digital content identifiers collected in the database; and

transmitting a content identifier to the client agent indicating the presence or absence of a characteristic in the file.

29. The method of claim 27 wherein said step of collecting comprises collecting a digital identifier for a data file.

30. The method of claim 27 wherein said file content is an e-mail.

9

A3

~~31. The method of claim 28 wherein said step of characterizing comprises:~~
~~tracking the frequency of the collection of a particular identifier;~~
~~characterizing the data file based on said frequency;~~
~~storing the characterization; and~~
~~comparing collected identifiers to the known characterization.~~

9